

CSA1618 DWDM-DE

EXPERIMENT-27

PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORITHM THROUGH WEKA

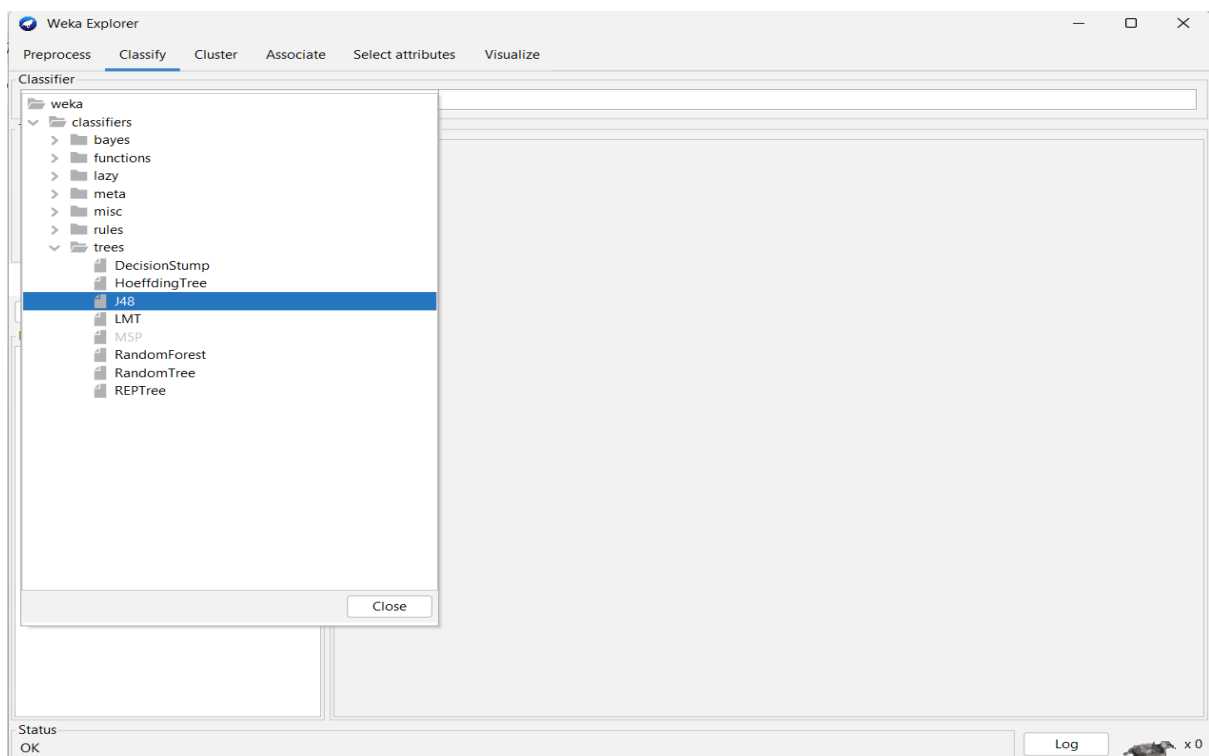
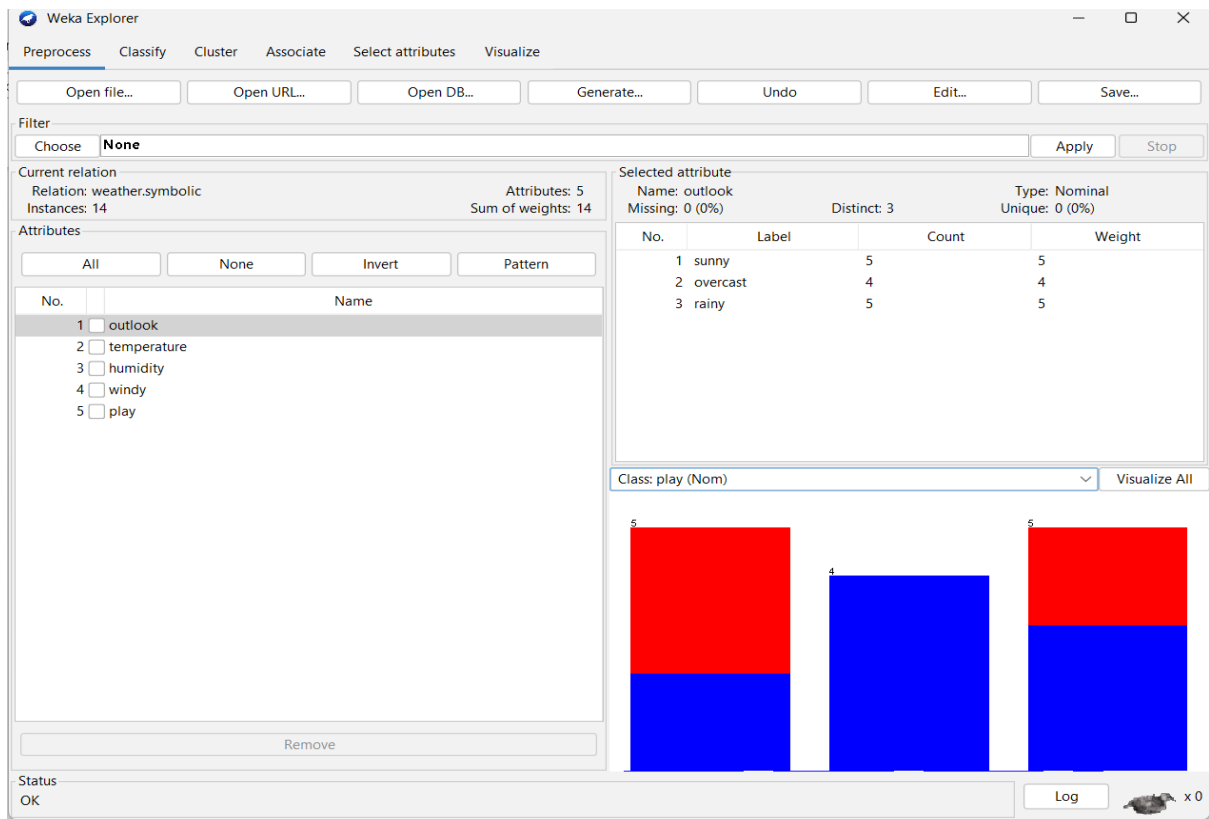
AIM:

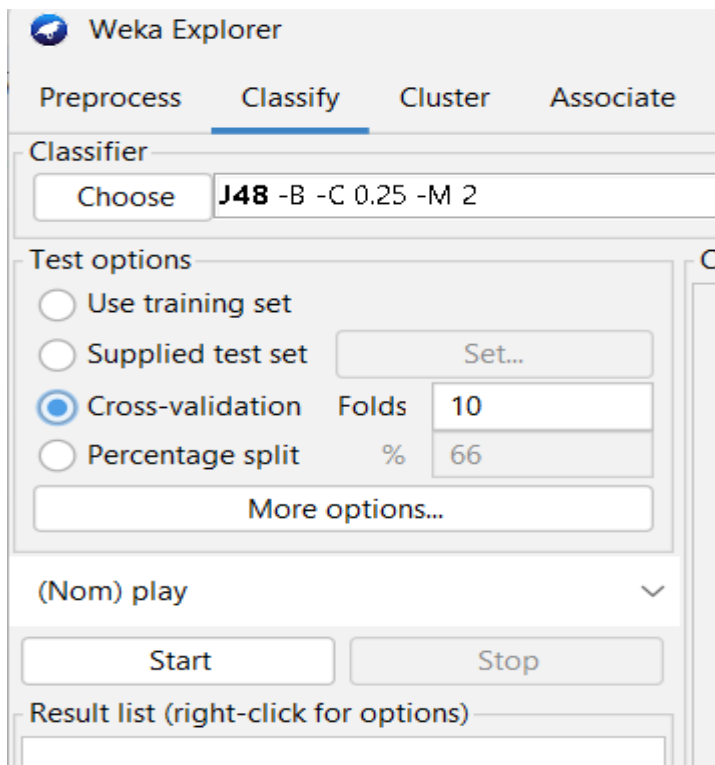
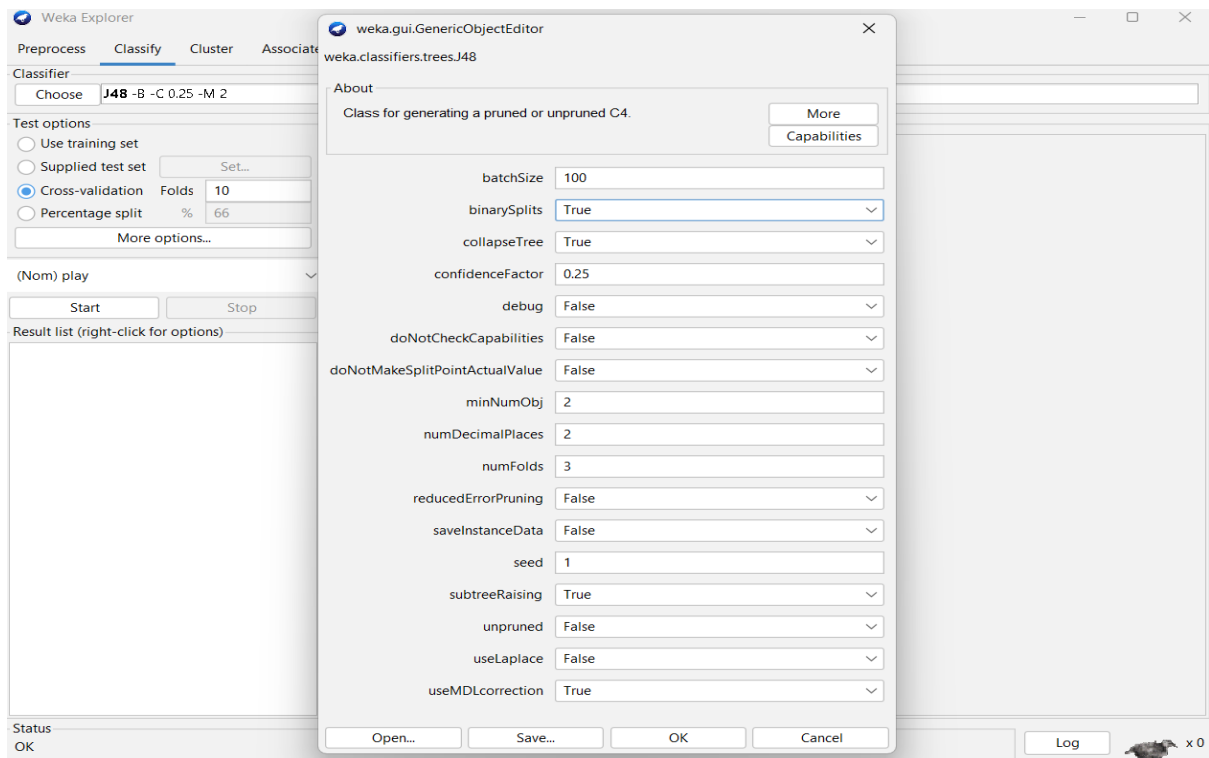
To create prediction of categorical data using decision tree algorithm through weka tool.

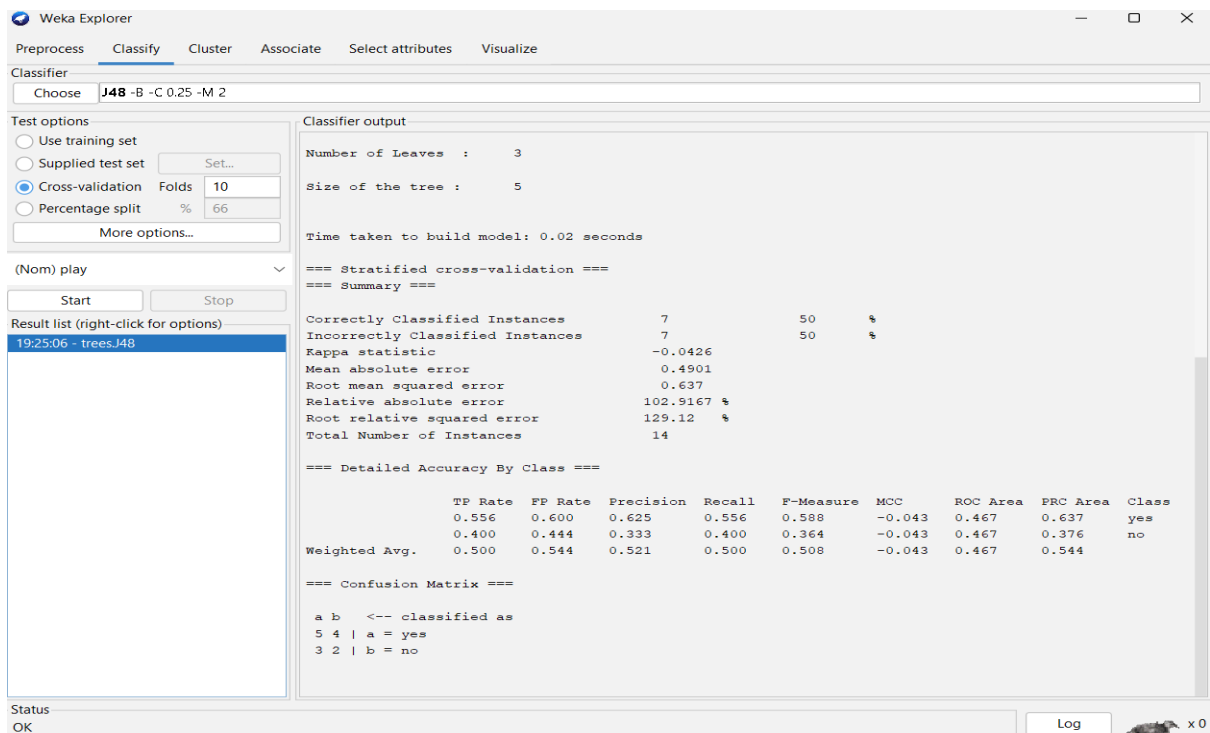
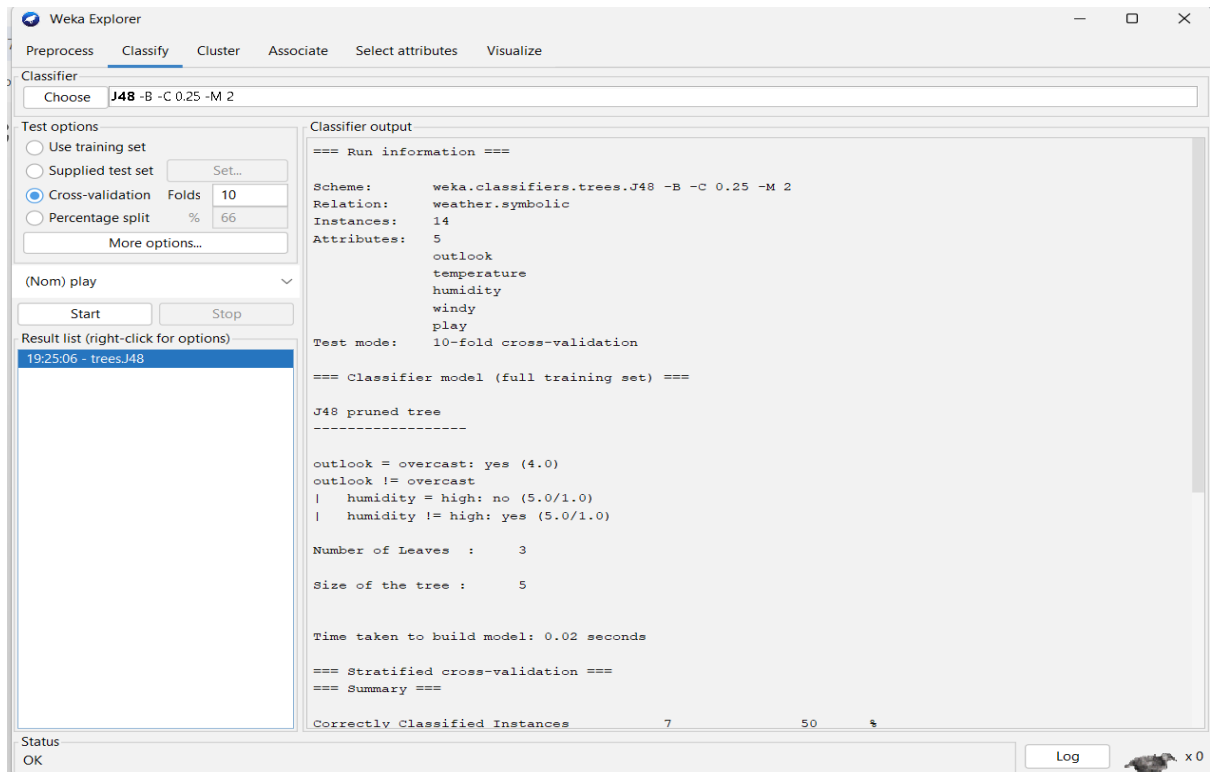
PROCEDURE:

1. Download and install WEKA.
2. Open WEKA and Choose "Explorer" from the main menu.
3. Under Preprocess, Click on the open file button and select the dataset. Ensure that your dataset contains categorical (nominal) attributes.
4. Go to the **Classify** tab, Click **Choose** → Expand the **trees** section → Select **J48** (C4.5 algorithm).
5. Click on J48 to open parameter settings. Modify parameters if needed: Confidence Factor (C): Default is 0.25 (lower values lead to more pruning), MinNumObj (M): Minimum number of instances per leaf (default is 2), Use binary splits (B): Set to True if needed.
6. Under "Test options", the following: Use training set: Trains and tests on the same dataset (not recommended), Cross-validation (default: 10-fold): Splits data into 10 parts for training/testing.
7. Click Start to begin classification.



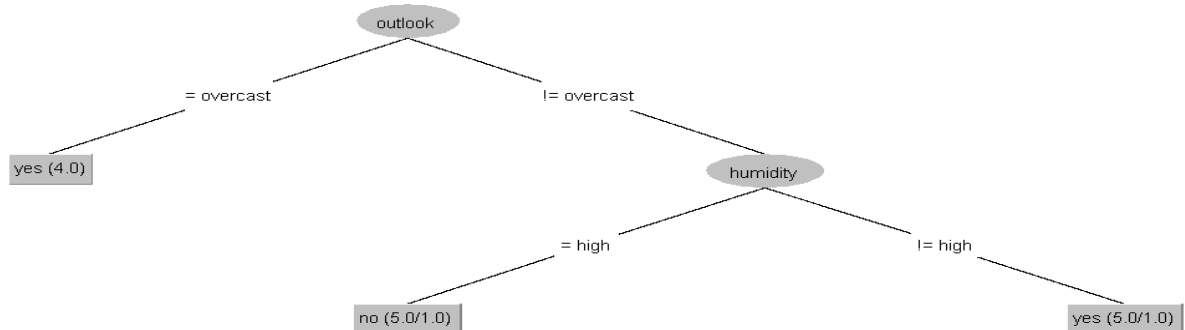






OBSERVATION:

Cross Validation folds:10



=== Confusion Matrix ===

a b <-- classified as

5 4 | a = yes

3 2 | b = no

Root relative squared error 129.12 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
yes	0.556	0.600	0.625	0.556	0.588	-0.043	0.467	0.637
no	0.400	0.444	0.333	0.400	0.364	-0.043	0.467	0.376
Wt Avg.	0.500	0.544	0.521	0.500	0.508	-0.043	0.467	0.544

Cross Validation folds:5

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
yes	0.444	0.600	0.571	0.444	0.500	-0.149	0.400	0.617
no	0.400	0.556	0.286	0.400	0.333	-0.149	0.400	0.337
Weighted Avg.	0.429	0.584	0.469	0.429	0.440	-0.149	0.400	0.517

=== Confusion Matrix ===

a b <-- classified as

4 5 | a = yes

3 2 | b = no

RESULT:

Thus, the observations and evaluations done on the dataset are analyzed. The decision tree has been successfully visualized. Various evaluations and comparisons done through the cross validation folds change. Which lead to the change of values in confusion matrix.